Learning a function

As mentioned earlier, we can generally think of a machine learning *algorithm* as a process for learning, and *models* as specific representations that we train using data. In essence, machine learning algorithms aim to learn a target function (f) that describes the mapping between data input variables (X) and an output variable (Y).

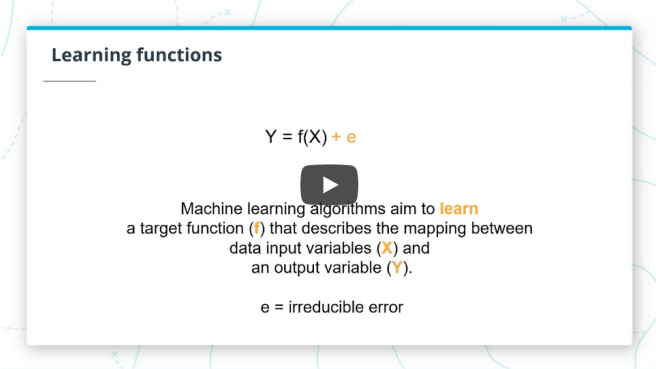
$$Y = f(X)$$

The core goal is to learn a useful transformation of the input data that gets us closer to the expected output.

Since the process extrapolates from a limited set of values, there will always be an error e which is independent of the input data (X) such that:

$$Y = f(X) + e$$

The variable e is called irreducible error because no matter how good we get at estimating the target function (f), we cannot reduce this error.



Note that the **irreducible error** we're discussing here is different from the **model error** we talked about earlier in the lesson. Irreducible error is caused by the data collection process—such as when we don't have enough data or don't have enough data features. In contrast, the model error measures how much the prediction made by the model is different from the true output. The model error is generated from the model and can be reduced during the model learning process.

QUESTION 1 OF 2

The goal of machine learning algorithms can be represented as a *learning function*:

$$Y = f(X) + e$$

What does each part of this represent?

Submit to check your answer choices!

PART	DESCRIPTION
Y	output variable
X	input variables
f	the target function
e	irreducible error

SUBMIT

QUESTION 2 OF 2

Which of the following statements is most accurate about machine learning?

- Learning a function from data is usually easy, since the models produced by modern algorithms have zero error.
- It is usually clear in advance which algorithm will produce the best function.
- ✓ In practice, it is often best to try a variety of algorithms and compare the results to see which produces the function with the least error.

SUBMIT

NEXT